

Membrane-Based Functions in the Origin of Cellular Life

Final Technical Report for Contract NCC2-1126

Michael A. Wilson

Dept. Pharmaceutical Chemistry, University of California, San Francisco, 94143

Period of Report: June 1, 1999 – December 31, 2002

Abstract

How simple membrane peptides performed such essential protocellular functions as transport of ions and organic matter across membranes separating the interior of the cell from the environment, capture and utilization of energy, and transduction of environmental signals, is a key question in protobiological evolution. On the basis of detailed, molecular-level computer simulations we investigate how these peptides insert into membranes, self-assemble into higher-order structures and acquire functions. We have studied the insertion of an α -helical peptide containing leucine (L) and serine (S) of the form (LSLLSL)₃ into a model membrane. The transmembrane state is metastable, and approximately 15 kcal mol⁻¹ is required to insert the peptide into the membrane. Investigations of dimers formed by (LSLLSL)₃ and glycoporphin A demonstrate how the favorable free energy of helix association can offset the unfavorable free energy of insertion, leading to self-assembly of peptide helices in the membrane. An example of a self-assembled structure is the tetrameric transmembrane pore of the influenza virus M2 protein, which is an efficient and selective voltage-gated proton channel. Our simulations explain the gating mechanism and provide guidelines how to re-engineer the channel to act as a simple proton pump. In general, emergence of integral membrane proteins appears to be quite feasible and may be easier to envision than the emergence of water-soluble proteins.

Introduction

Our research is devoted to the origin of cellular functions, with a long-term objective to explain how protocells performed functions essential for their survival and evolution utilizing only the molecules that may have been available in the protobiological milieu. We have developed simple, molecular models of several protocellular functions, and examine their structure, stability and mechanism of action.

Hypotheses. A basic hypothesis underlying much of this work is that boundary structures made of membrane-forming material emerged early in the course of protobiological evolution and served as precursors to protocells. We further assume that protocellular structures evolved into contemporary functional units without undergoing discontinuous transitions (Morowitz *et al.*, 1988). Our concepts of protocellular functions are therefore based on our knowledge of contemporary cells. One consequence of this approach is that, at some early stage of protocellular evolution, the primary metabolic functions were performed by peptides, the precursors of the proteins in modern cells. We postulate that protobiologically plausible models of these functions share structural and mechanistic motifs with contemporary proteins.

Our hypotheses do not imply that peptides were the first functional molecules in protocells. Other molecules may have preceded or co-existed with peptides. In this sense, the proposed work does not depend on a specific scenario for the origin of life. We also do not explicitly address the

question about the origin of peptides or their building blocks, amino acids. We simply assume that their emergence was necessary on the evolutionary pathway to proteins.

Protocells and their functions. Probably the first cell-like structures were vesicles — closed, spheroidal assemblies of organic material enclosing an aqueous medium. The walls of vesicles are built of amphiphilic molecules which have water-soluble (hydrophilic) and water-insoluble (hydrophobic) groups at opposite ends. These molecules are arranged in bilayers such that the hydrophilic head groups point toward water and the hydrophobic tails form the interior of the bilayer. In this respect, vesicle walls resemble modern cell membranes. Under the proper conditions, vesicles form spontaneously from an aqueous solution of amphiphiles. The source of amphiphiles on the primitive earth might have been terrestrial or extraterrestrial and, in fact, vesicles form from the material extracted from the Murchison meteorite (Deamer and Pashley, 1989).

Vesicles became the precursors to true cells — protocells — by acquiring the capabilities needed to survive and reproduce. Protocells had to transport ions and organic matter from the environment across their walls, capture and utilize energy, and synthesize the molecules necessary for self-maintenance and growth (Morowitz *et al.*, 1988). The identity of molecules that performed these functions is open to debate. Short polymers of RNA are attractive candidates because they could act as both catalysts and information storage systems. Since even small RNA molecules can maintain a rigid three-dimensional structure, they are well suited to act as proto-enzymes (Joyce, 1996). In fact, several simple RNA enzymes have already been created in the laboratory (Hager *et al.*, 1996). However, the concept of RNA molecules as the sole functional species in protocells encounters difficulties. RNA is fragile, easily hydrolyzed in water, and no efficient prebiotic syntheses of its building blocks have been found. Furthermore, it has not been shown that RNA can be incorporated into membranes to perform functions that, in modern cells, include energy transduction and transport.

In modern organisms, most metabolic functions are carried out by proteins. The most parsimonious assumption is that their protobiological precursors were peptides. Their protocellular potential is illuminated by the fact that a wide range of simple, naturally occurring (Cafiso, 1994) or synthetic (Lear *et al.*, 1997) peptides can spontaneously insert into membranes and assemble into channels capable of transporting material across cell walls. In protocells, simple, functional peptides may have emerged independently of possible RNA enzymes or may have been synthesized by these enzymes. In either scenario, they constituted an essential step on the pathway to the cellular metabolism as we know it now.

Primitive catalysis. One of the main functions of a protocell was to catalyze the chemical reactions needed for its metabolism. In nature, this role is filled almost exclusively by protein enzymes. It is natural to expect that they evolved from peptide catalysts at some stage of protocellular evolution. Unfortunately, efforts to design peptide catalysts have been mostly unsuccessful (Corey and Corey, 1996).

The essential features of enzymatic catalysis are the formation of substrate-enzyme associations accompanied by the entropic effects of substrate immobilization, the exclusion of solvent from the active site, the presence of specific chemical groups at fixed locations in the enzyme and the preferential stabilization of the transition state complex. All of these features are presently difficult to achieve with small, flexible, water-soluble peptides. An instructive example is an attempt to build, *de novo*, a water-soluble peptide that mimics a serine protease, chymotrypsin (Hahn *et al.*, 1990). Considerable efforts were made to design a structure that incorporated both the canonical catalytic triad of amino acids and the oxyanion hole (binding site) of this enzyme, but the initially reported rate of 0.1-1% of the rate of the native enzyme was later found to be erroneous (Corey and Corey, 1996). The peptide failed to exhibit any of the features listed above and, consequently,

showed no catalytic activity.

There are also some positive examples (Johnsson *et al.*, 1993; Brack and Barbier, 1990; Perez-Paya *et al.*, 1994; Severin *et al.*, 1997). Their common feature is a well defined geometry of the catalyst interacting with the substrate. In at least two cases, properties of surfaces were exploited (Brack and Barbier, 1990; Perez-Paya *et al.*, 1994), indicating that this is a promising direction in search for simple, peptide catalysts.

Even more promising are the recent techniques for *in vitro* evolution of functional peptides (Roberts and Szostak, 1997). These techniques should eventually bring the wealth of information about the catalytic potential of peptides. It may be suggested that they will make computer simulations, based on rational design, unnecessary. We feel that the opposite is true; understanding the structure of new catalysts will become an important research area leading to further improvements of *in vitro* experiments.

Protocellular bioenergetics. To perform their metabolic functions, protocells had to transduce energy captured from the environment. In contemporary cells, this is accomplished by a variety of complex mechanisms. Most of them, however, share the common feature that the acquired energy is converted into a transmembrane proton gradient (Skulachev, 1984) used for the synthesis of "high-energy" compounds. The universality of this mechanism suggests that it arose early in protobiological development.

Different early environmental energy sources have been proposed, including chemical energy and light (Deamer, 1997). Probably the simplest system capable of converting light into transmembrane proton gradients consists of polycyclic aromatic hydrocarbons (PAHs) incorporated into vesicle membranes. (Deamer, 1992) Upon photo-excitation, the PAHs release protons either to the exterior or the interior of the liposome. Protons in the environment dissipate while those inside the liposome accumulate, thereby creating a proton gradient. Since protons in the interior of the liposome are also used to regenerate the initial state of the system, this proton gradient is only transient. What is needed is a "gate-keeper" mechanism to ensure that reprotonation of the proton source does not dissipate the already formed proton gradient. Another simple, light-activated proton pump that incorporates some directionality has been recently constructed by exploiting electrical properties of membrane surfaces. (Sun and Mauzerall, 1996)

The best understood, contemporary, biological system for photo-generating transmembrane proton gradients is bacteriorhodopsin, a membrane protein in the microorganism *Halobacterium salinarium*. (Lanyi, 1997) In bacteriorhodopsin, light energy is absorbed by a retinal chromophore attached to the protein near the center of the bilayer. The excited retinal undergoes a conformational change which results in the release of a proton to a nearby amino acid side chain in the protein. This, in turn, changes the electrostatic environment of the surrounding amino acids and facilitates further proton transfer along the protein, possibly assisted by transient water bridges, until the proton is ejected into aqueous solution. The accompanying relaxation of the retinal conformation prevents its reprotonation transferred proton.

Although it is unlikely that bacteriorhodopsin was the earliest proton pump, the mechanism of its action illustrates a robust principle. Transport or separation of charges through proton transfer between side chains of a protein in response to the changes in local environments appear to be ubiquitous in biology. It is observed not only in membrane proteins, but also in water-soluble enzymes, of which proteases are probably the best known example. Thus, it is only natural to invoke a similar mechanism in models of ancient proton pumps.

Modern cells utilize the energy stored in the transmembrane proton gradient to drive the synthesis of high-energy chemical bonds, such as adenosine triphosphate (ATP). For example, in enzymes called ATP synthases the dissipative flow of protons through the enzyme is coupled to the synthesis

of ATP. Although other energy storage systems, such as thioester bonds, may have preceded ATP this example illustrates that both *active and passive* proton transport across cell walls must be understood in order to develop protocellular models for precursors of modern bioenergetic systems.

A host of naturally occurring and synthetic proteins and peptides are capable of transporting charges across membranes. One example is gramicidin A, a (D,L) peptide whose dimers form membrane-spanning pores that allow the permeation of protons and monovalent cations (Wooley and Wallace, 1992). Most membrane channels, however, are formed by the self-assembly of proteins such that several (typically four or five) α -helices are arranged around a permeating pore. Some exceedingly simple peptides were shown to form such ion channels (Oliver and Deamer, 1994; Lear *et al.*, 1997). These structural similarities suggest that α -helical assemblies of transmembrane peptides lie at the origin of charge-transporting systems. *in vitro* experiments.

Methods

The best approach to the computational study of biological systems at a molecular level has been the molecular dynamics method. Here, we outline the basic aspects of this method that are relevant to the simulations described in this paper. Several books (Allen and Tildesley, 1987; Frenkel and Smit, 2001) provide exhaustive descriptions of molecular dynamics, as applied to chemical and biological systems.

In molecular dynamics, Newton's equations of motion are solved numerically for all of the atoms in the system under study using an iterative procedure. From the positions, velocities and forces acting on the atoms at time t , new positions and velocities at time $t + \delta t$ can be calculated if the forces do not change appreciably over a time step, δt . A typical time step is equal to 1 femtosecond (10^{-15} s). By repeating this procedure many times we obtain a time-history of the system, called a trajectory. State-of-the-art molecular dynamics trajectories for biological systems extend from 1 nanosecond to 1 microsecond (10^{-9} s – 10^{-6} s), which require generating 10^6 – 10^9 time steps.

Potential Energy Functions. Solution of Newton's equations of motion requires knowledge of the forces acting on the atoms in the system. These forces depend on interactions between the atoms and are computed from a potential energy function, which describes the stretching of chemical bonds, the bending of valence angles, the rotation of dihedral angles and the electrostatic and van der Waals interactions between atoms. Much work has been invested in the construction of potential energy functions that successfully reproduce properties of water-soluble proteins, (Case *et al.*, 1999; MacKerell *et al.*, 1998) pure membranes (Berger *et al.*, 1997; Smondyrev and Berkowitz, 1999; Feller, 2000) and their mixtures with small solutes and peptides. (Bassolino-Klimas *et al.*, 1995; Damodaran *et al.*, 1995; Shen *et al.*, 1997)

We used the TIP4P potential model for water interactions, (Jorgensen *et al.*, 1983) which reproduces many of the thermodynamic properties of liquid water. The octane molecules were treated at the united-atom level using the OPLS potentials (Jorgensen *et al.*, 1984). These potential models have been shown to provide a good description of the thermodynamic and structural properties of pure water and alkane phases as well as the water-alkane interface. The potentials used to describe the interactions between phospholipids have already been successfully applied in simulations of bilayer systems. (Essmann *et al.*, 1995a; Chiu *et al.*, 1995; Tieleman *et al.*, 1997) The protein molecules were treated at a full atomic level of detail using the AMBER force field (Case *et al.*, 1999).

In many instances, considerable savings of computer time without a significant sacrifice of accuracy can be achieved by smoothly truncating interactions between pairs of atoms in the system at a specified distance, typically 8-10 Å. However, phospholipid head groups and ionizable side chains

of amino acids carry large charges, which interact non-negligibly over long distances, extending beyond the primary simulation cell. In these instances, truncation of interatomic interactions may be insufficiently accurate. The most common method for calculating long-ranged effects is called Particle-Mesh Ewald (PME) (Essmann *et al.*, 1995b). In PME, the long-range, electrostatic interactions are evaluated through the solution of a differential equation on a grid, using the Fast Fourier Transform (FFT) method. We have employed this approach in our simulations of transmembrane channels in phospholipids. (Schweighofer and Pohorille, 2000)

Description of the systems. The simulated systems consisted of a model peptide either embedded in a bilayer solvated by water or located at a model water-membrane interface. All components of the system were represented at the atomic level. The system was placed in a box — the primary simulation cell. The number of atoms in the system is typically in the 10^4 – 10^5 range. The corresponding cross-sectional length of the water-membrane interface varied between 4 and 6 nm. To obtain results for a macroscopic system from such simulations, the content of the primary cell was periodically replicated in space, forming an infinite lattice of identical simulation cells. This is a standard approach to removing edge effects in molecular simulations.

Many properties of the system can be computed directly from a molecular dynamics simulation trajectory. Thermodynamic properties, such as the temperature, pressure, or membrane surface tension, can be expressed as averages over the series of configurations that form the trajectory. Structural quantities, such as conformational parameters characterizing the protein or its individual residues can be computed in the same fashion. All these quantities can be directly compared with the same quantities measured experimentally.

Free energy calculations. Often, a goal of computer simulations is to determine how the free energy of the system changes in the course of a chemical or biochemical process. These changes are directly related to the relative stabilities of different states of the system. Since free energies cannot be expressed as statistical averages of mechanical properties, special techniques are required for their evaluation. (Frenkel and Smit, 2001; Berne and Straub, 1997) Two of these techniques were used in the simulations discussed in the subsequent sections.

In one approach, a series of simulations is performed, in which the system is constrained to several, overlapping ranges, or “windows”, along an appropriately chosen, physical degree of freedom, often called “the reaction coordinate”, ξ . For example, to compute the free energy change accompanying the insertion of a peptide into the membrane, the distance between the center of mass of the peptide and the midplane of the lipid bilayer could be defined as such a coordinate. For each window, the probability, $\mathcal{P}(\xi)$, of finding the system at different values of the chosen coordinate is obtained. This probability defines the change of the free energy along ξ , $\Delta A(\xi)$, through the relation:

$$\Delta A(\xi) = -k_B T \log \mathcal{P}(\xi) \quad (1)$$

where k_B is the Boltzmann constant and T is the temperature of the system. $\Delta A(\xi)$ over all windows is obtained from the requirement that it must be a continuous function of the chosen coordinate.

Another method for estimating free energy changes associated with the point mutations of a given amino acid in a peptide, or with evolution of the system along a “reaction coordinate”, is based on the free energy perturbation method (Zwanzig, 1954). For point mutations, this method is implemented *via* an “alchemical transformation”, in which the residues of interest in the wild type are perturbed into those of the mutant. In practice, residues are altered by modifying separately their point charges, van der Waals parameters and internal coordinates — *i.e.* shrinking or growing

chemical bonds — as a function of a coupling parameter. (Kollman, 1993) In the simulations described in this paper, creation or annihilation of non-bonded and internal parameters were carried out using the single topology approach, thus eliminating the need for defining distinct topologies for both the initial and the final states of the mutation. (Pearlman, 1994)

Results

Short peptides tend to accumulate at interfaces and acquire ordered structures, provided that they have a proper sequence of polar and nonpolar residues. The specific identity of the amino acids appears to be less important for this process, a desirable protobiological property. The driving force that enables or enhances secondary structure formation for proteins interacting with or incorporated into membranes is the hydrophobic effect, which is manifested at aqueous interfaces as a tendency for polar and nonpolar groups of the solute to segregate into the aqueous and nonpolar phases, respectively. The emerging amphipathic structures are strongly favored.

Among these structures, α -helices are especially stable because they are further stabilized by intramolecular hydrogen bonding interactions. In bulk water, these interactions do not contribute to the stability of helices because of competing interactions between hydrogen bonding centers and water molecules.

If peptides consist of nonpolar residues only, they become inserted into the nonpolar phase. As demonstrated by the example of the L-leucine undecamer, nonpolar peptides tend to fold into an α -helix as they partition into the nonpolar medium. Once in the nonpolar environment, the peptides can readily change their orientation with respect to the interface from parallel to perpendicular, for example in response to local electric fields. (Tieleman *et al.*, 2001) The ability of nonpolar peptides to respond to changes in external conditions may have provided a simple mechanism for transmission of signals from the environment to the interior of a protocell.

Insertion of Peptides into Membranes. According to the two-state model, interfacial folding of transmembrane proteins is followed by their insertion into the bilayer. A transmembrane α -helical peptide must contain approximately 20 residues to extend for the full width of a membrane that is 3 nm thick. We have studied one such peptide built of only two amino acids — L-leucine and L-serine (S). The peptide contains 21 residues in the sequence (LSLLLSL)₃. This sequence has been chosen such that in the α -helical form the peptide is amphipathic, *i.e.* all serine residues, which are polar, lie along the same face of the α -helix. Despite its simplicity, (LSLLLSL)₃ exhibits several interesting properties. It was shown experimentally that the peptide formed transmembrane, tetrameric ion channels in the presence of an electric field. (Lear *et al.*, 1988) When the electric field was removed, the channels persisted on time scales of milliseconds before the individual peptides reverted to their resting state parallel to the water-membrane interface, indicating that the transmembrane channels do not correspond to the global free energy minimum of the peptide, but are weakly metastable.

Experimental results, however, provide no information about stability of individual helices in the transmembrane orientation. This depends on the balance of hydrophobic forces, which tend to drive the nonpolar leucine residues into the membrane interior, the hydrophilic forces, which are favorable when the serine residues are located in the aqueous solution, and the interactions of unsaturated hydrogen bonding sites at both ends of the α -helix with the water. The transmembrane and in-plane states are shown in Fig. 1.

The peptide in the in-plane state, oriented parallel to the interface, remains α -helical. The hydrophilic face, composed of 6 serine residues, points to the water where the hydroxyl groups of the serine residues can hydrogen bond with the water. The hydrophobic face points into the octane. This can be clearly seen in Fig. 2 which shows the density profiles of the C $_{\alpha}$ carbon atoms

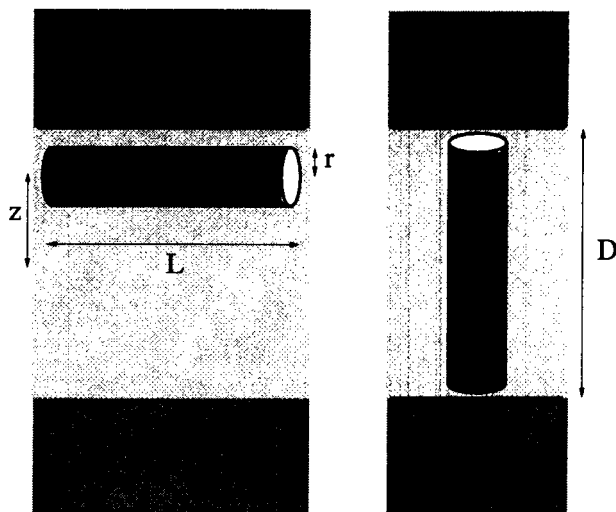


Figure 1: Schematic picture of a helical peptide near a water-oil (membrane mimetic) interface in orientations parallel (left) and perpendicular (right) to the interface. The width of the lamella (or membrane) is D , the length of the peptide helix is L and the radius is r . The distance between the center of the membrane and the center-of-mass of the peptide is z .

of the serine and leucine residues. The center-of-mass of the LS peptide is located at $z = 14$ Å from the center of the octane lamella, which means that most of the molecule is on the oil side of the interface. This is to be expected due to the smaller free energy required to create cavities that are large enough to contain the peptide in octane and water (Pohorille and Wilson, 1996).

While the hydrogen bonding interaction between the serine residues and water make a significant contribution to the surface activity of the peptide, the N-acetyl and N'-methyl blocking groups at the ends of the peptide also interact strongly with water. Simulations of a blocked undecamer of poly-leucine at a water-hexane interface demonstrated that the ends of the peptide interact quite strongly with the aqueous phase, and there is a substantial free energy barrier to removed the poly-leucine from the interface into the membrane interior (Chipot and Pohorille, 1998). Of course, the poly-leucine does exhibit much greater orientational flexibility, and orientations of the helix axis almost perpendicular to the interface were observed. The interaction of the serine residues with the water inhibit this type of motion, and the peptide was not observed to spontaneously rotate out of the plane of the membrane over simulation times of 10 ns.

To clarify the issue of helix stability, we calculated the free energy of inserting the peptide into a model membrane system using a variant of the free energy window method, described briefly in the methods section. Fig. 3 shows the free energy of the peptide as a function of the location of its center-of-mass relative to the center of the membrane ($z = 0$). The water-membrane interface is located at $z = -13$. The orientation of the peptide depends on the location of its center-of-mass. At $z = 0$ the peptide is approximately perpendicular to the plane of the membrane, with an end in each of the aqueous phases (a transmembrane orientation). In contrast, at $z = 13$, the peptide lies parallel to the interface such that the serine residues point towards the water. The two main features of the curve in Fig. 3 are that the in-plane state is approximately 20 kcal mol^{-1} more stable than the transmembrane state and that the latter state corresponds to a broad and shallow free energy minimum. This means that the transmembrane state of the peptide is, at best, only weakly metastable and a single peptide in this state will quickly convert to the in-plane state in

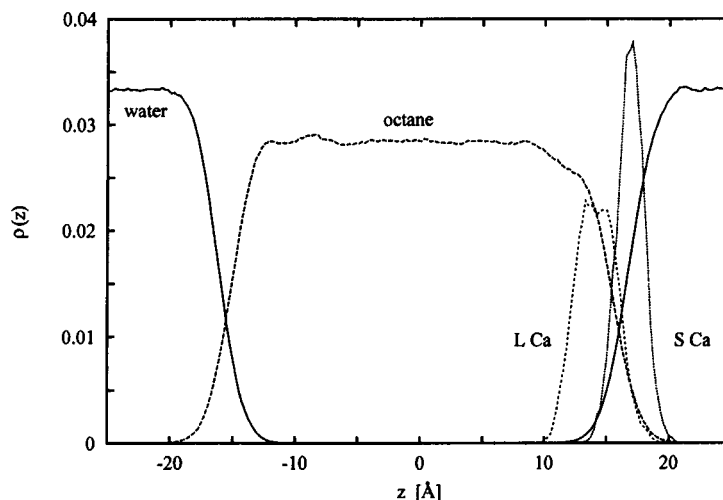


Figure 2: Density profiles of water (red), octane (green), C_{α} carbons of serine (magenta) and of leucine (Blue) when the LS peptide is oriented parallel to the water-octane interface.

the absence of an electric field.

These features of the free energy surface were borne out by independent simulations with the peptide initially located in either an in-plane or transmembrane orientation. The in-plane state was stable over the course of a 10 ns simulation. The peptide backbone remained entirely α -helical and the serine residues always pointed towards the water. The transmembrane peptide adopted a variety of mixed α -helical and 3_{10} -helical arrangements. This is due to a mismatch between the length of the peptide and the width of the membrane. The peptide in the α -helical conformation is slightly longer than the width of the membrane. Converting part of the backbone to 3_{10} lengthens the helix and allows for the formation of additional, energetically favorable, serine-water hydrogen bonds. In simulations of a system with a somewhat thicker hydrophobic membrane core, the peptide remained α -helical.

A total of three trajectories were started with the peptide in the transmembrane state. On the basis of the free energy calculations it was expected that this state would not be stable and, over time, the peptide would move to the water-membrane interface. In two of the simulations, the peptide spontaneously converted from the transmembrane to the in-plane state after 7 and 9 ns, respectively. In the third simulation, however, the peptide remained transmembrane after 18 ns. This is due to the asymmetry in the ends of the peptide. As found for the undecamer of poly-L-leucine, the C-terminus interacts with the water much more strongly than the N-terminus. The flat free energy curve in the region near $z = 0$ indicates that the peptide can readily diffuse towards either of the two water-membrane interfaces. In the two simulations, in which the peptide converted to the in-plane state, it initially diffused in the direction that required dehydration of the N-terminus. Since this end interacts with water relatively weakly the conversion appears to proceed quickly with little or no free energy barrier. In contrast, in the third simulations, the peptide initially diffused in the opposite direction, which would have required dehydration of the C-terminus. Considering highly favorable interaction of this terminus with water such dehydration process is unlikely. Instead, it is expected that in a sufficiently long molecular dynamics trajectory the peptide would diffuse back to the center of the bilayer and, eventually, converted to the in-plane state by dehydrating its N-terminus.

In the simulations of the metastable, transmembrane monomer, the helix axis quickly adopted a tilt angle of about 30 degrees. This large amount of tilt was accompanied by a shift away from

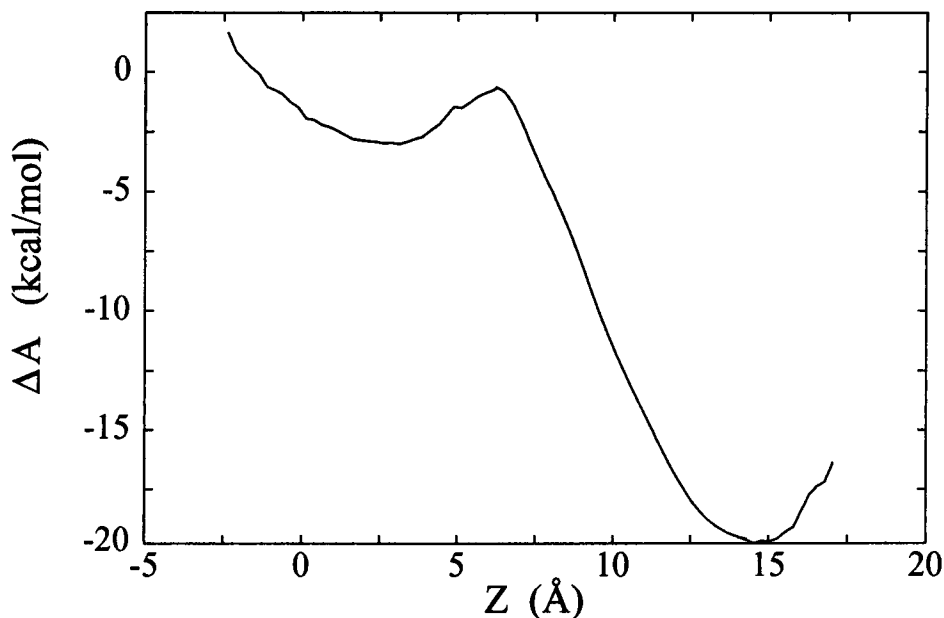


Figure 3: Free energy profile of the center-of-mass of the (LSLLSL)₃ peptide as a function of its position relative to the center of the model membrane. The center of the membrane is located at $z = 0$ Å and the water membrane interface is located at $z = 15$ Å.

the purely α -helical starting configuration to a mixture of α and 3_{10} helices. The 3_{10} helix is more tightly wound than the α -helix, with 3.0 residues per turn as compared to 3.6 residues per turn in an α -helix. This is due to two effects. First, the tighter wrap of the helix leads to an overall lengthening of the molecule, which allows more serine residues to interact with the water surfaces. Second, the tighter helix means that the hydrophilic residues no longer lie along a face of the helix, but twist around the helix. This also increases the interaction of the hydrophilic groups with both interfaces.

Helix Association in Membranes

Single transmembrane helices are rarely capable of performing biological functions. Instead, they form functional units after self-assembling into higher order structures. However, not all helices self-assemble. Consequently, it is necessary to understand sequence-specific interhelical recognition before we can predict the kinds of structures that could have formed in protocellular membranes. The simplest models for peptide association are helical dimers. Although they cannot form channels, some are biologically active. Moreover, it is assumed that dimer formation is the first step in aggregation into higher order assemblies. For example, it has been suggested that tetrameric channels are formed as "dimers of dimers". (Zhong *et al.*, 1998b)

Given the existence of peptide association, we note that peptides can either be inserted into the membrane where they associate into dimers or larger multimeric structures or associate at the interface and only then become inserted into the membrane. The second scenario might seem plausible, especially for amphipathic peptides because their aggregates could be stabilized by interhelical hydrogen bonds between polar residues. The nonpolar residues of the aggregate would be exposed to the aqueous environment, which should promote insertion into the nonpolar membrane.

To test this mechanism of aggregation, we have carried out simulation studies on dimers of the (LSLLSL)₃ peptide, which has already been discussed in the previous section. When the dimer

was placed parallel to the plane of the membrane at the water-membrane interface, it dissociated in less than 2 ns. The interfacial water molecules successfully competed for the serine hydrogen bonding sites, which led to the loss of serine-serine hydrogen bonds. Additionally, interactions between electrical dipoles associated with the α -helices are highly unfavorable in a parallel arrangement. These results indicate that self-assembly of peptides at interfaces is unlikely. In contrast, a transmembrane dimer was found to be stable over the course of a 15 ns simulation. Near the ends of the dimer, serine-serine hydrogen bonds were lost in favor of water-serine hydrogen bonds, allowing water molecules to penetrate the membrane around the peptide. This, in turn, might increase rates of non-specific permeation of ions and polar solutes across membranes. (Deamer and Nichols, 1989; Paula *et al.*, 1996; Paula *et al.*, 1998; Wilson and Pohorille, 1996; Pohorille and Wilson, 2001) However, the serine-serine hydrogen bonds in the middle of the dimer remained intact, keeping the helices in the dimer together. These results confirm that association of two peptides appears to increase the stability of the transmembrane state relative to isolated monomers.

Functions of Membrane Peptides — A Model Transmembrane Proton Transport System

Aggregates of membrane proteins are of special interest if they can perform important cellular functions. One such function is transport of protons across membranes, which is an essential process for both bioenergetics of modern cells and the origins of cellular life. All living systems convert environmental energy into chemical energy by using transmembrane proton gradients to drive the synthesis of adenosine triphosphate (ATP) from adenosine diphosphate (ADP). ATP, in turn, is used as a source of energy to drive many cellular reactions. The ubiquity of this process in biology suggests that even the earliest cellular systems relied on proton gradients to harvest the energy needed for their survival and growth. In contemporary cells, proton transfer is assisted by large, complex proteins embedded in membranes. Could the same process have been accomplished with the aid of similar, but much simpler peptides that could have existed in the protobiological milieu?

To answer this question it is desirable to have a protein model which is small, has a well known structural motif, yet which operates with the efficiency and control of more complex proteins. This led us to study the Influenza-A M_2 protein, which forms small, voltage-gated proton channels. (Pinto *et al.*, 1992; Wang *et al.*, 1993; Sakaguchi *et al.*, 1997; Pinto *et al.*, 1997; Zhong *et al.*, 1998a; Forrest and Sansom, 2000; Lin and Schroeder, 2001; Mould *et al.*, 2000) The M_2 protein contains 97 amino acids, including a single transmembrane domain 19 residues long. Not all residues, however, are essential for transport. Active channels have been reconstituted from a synthetic peptide containing a subset of only 25 amino acids, including the transmembrane region, with no loss in specificity or efficiency. (Duff and Ashley, 1992)

In lipid bilayers, four identical protein fragments, each folded into an α -helix, aggregate to form small channels spanning the membrane. Protons are conducted through a narrow pore in the middle of the channel. Compared with a well-studied, proton-permeating peptide, gramicidin A, the rate of proton transport across the truncated M_2 channel is over 1000-fold faster. Remarkably, in contrast to gramicidin A, the M_2 channel is virtually impermeable to alkali ions, such as Na^+ and K^+ . This combination of efficiency and specificity makes M_2 an excellent, simple model to study the formation of proton gradients across membranes.

The channel is large enough to contain water molecules and is normally filled with water. In analogy to the mechanism of proton transfer in some other channels, (Akeson and Deamer, 1990; Schumaker *et al.*, 2001) it has been postulated that protons are translocated along the network of properly aligned water molecules filling the pore. This mechanism, however, must involve an additional, important step because the channel contains four L-histidine (H) amino acid residues,

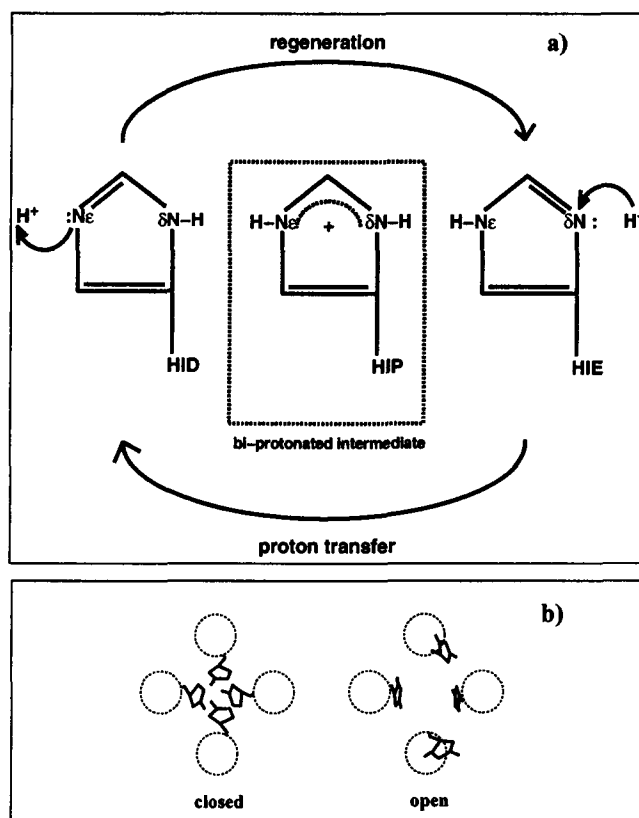


Figure 4: Schematics of two proposed models of gated proton transport in M2: (a) a proton shuttle mechanism mediated by the His³⁷ residues, where the initial state is regenerated by tautomerization and (b) a structural model where the gating is mediated by steric forces resulting from protonation of one or more His³⁷ residues.

one from each of the helices, which are sufficiently large to occlude the pore and interrupt the water network. The L-histidine residues have been implicated in gating protons. Due to their size, they ensure channel selectivity by blocking small ions, such as Na⁺ and K⁺, from permeating the membrane but provide a mechanism for proton transport. The role of the L-histidines in gating is supported by findings that point mutations, in which the L-histidines are substituted by other residues greatly impede the ability of M₂ to transport protons.

Two mechanisms of gating have been proposed, which rely on the ability of each L-histidine to become positively charged by accepting an additional proton. In one mechanism, all four L-histidines acquire a proton and, due to repulsion between their positive charges, move away from one another, thus opening the channel. (Sansom *et al.*, 1997) The alternative mechanism involves the ability of protons to move between different atoms in a molecule (tautomerization). In this mechanism, a proton is captured on one side of the gate while a second proton is released from the opposite side, and the molecule returns to the initial state through tautomerization. (Pinto *et al.*, 1997) These two mechanisms are shown schematically in Fig. 4.

Atomic-level molecular dynamics simulations were designed to test these two mechanisms. (Schweighofer and Pohorille, 2000; Pohorille *et al.*, 1999) The model system used in the study contained a bilayer membrane made of phospholipid, dimyristoylphosphatidylcholine (DMPC), which is a good model of the biological membranes forming cellular boundaries. Both sides of

the bilayer were surrounded by water which simulated the environment inside and outside the cell. Embedded in the membrane was a channel made of 25 amino acids fragments of the Influenza-A M₂ protein, and enough sodium counterions to maintain system neutrality. Several protonation states of L-histidine residues were considered. They represented different intermediate states of the channel predicted by the two proposed mechanisms of proton transport. The simulations revealed that all intermediate states of the system involved in the tautomerization mechanism were structurally stable and the arrangement of water molecules in the channel was conducive to the proton transport. In contrast, in the four-protonated state, postulated to exist in the gate-opening mechanism, the electrostatic repulsion between the L-histidine residues appeared to be so large that the channel lost its structural integrity and one helix moved away from the remaining three. These results indicate that translocation along a network of water molecules in the channel and tautomerization of the L-histidine residues is a likely mechanism of proton transport whereas a mechanism involving protonation of all four L-histidines is unlikely. A possibility of gate opening after protonation of two rather than four L-histidines has not been excluded.

These results not only explain how a simple protein system can achieve highly efficient and selective passive proton transport (i.e. transport along the concentration gradient) across cell walls, but also indicate how the system can be genetically re-engineered to become a simple directional, reversible proton pump. First, M₂ must be coupled with a chromophore capable of releasing a proton in response to light. Several very simple chromophores, such as polycyclic aromatic hydrocarbons, are already known. In fact, some of them have been shown to dissolve in membranes and generate transient, light-induced proton gradients. (Deamer, 1992) To maintain the proton gradient, it must be ensured that release of the pumped proton is followed by reprotonation of the chromophore with a proton from the opposite side of the membrane. This will involve manipulating the sequence of amino acids along the pore. Cysteine scanning mutagenesis has already shown that the replacement of the pore-lining, but not other residues can modify the properties of the channel. (Shuck *et al.*, 2000) Other designs of proton pumps are possible, based on coupling electron and proton transfer using iron or quinones. These have been recently shown to be of possible protobiological relevance. (Bernstein *et al.*, 2001) If such an experimental effort were to be successful, it would demonstrate that protein-based proton pumps could have emerged early in protobiological evolution. Furthermore, such a pump could be used to provide energy to laboratory-built models of protocells and cell-like structures built for biotechnological applications. (Lanyi and Pohorille, 2001)

Conclusions

Many proteins that perform essential cellular functions are embedded in membranes that encapsulate cells or cellular components. These proteins or protein complexes are among the largest macromolecular structures found in cells and their mode of action is often complicated and subtle. This appears to create a serious difficulty from the origin of life point of view. If the functions performed by membrane proteins are essential to the existence of even the simplest cells, how could they have been performed, even if less efficiently or selectively, by much simpler peptides?

Here, we have argued that the emergence of integral membrane proteins may have been quite feasible. In fact, this may be much easier to envision than the emergence of water-soluble proteins. We have supported our arguments with results of our molecular dynamics computer simulations and a considerable body of evidence from other experimental and theoretical studies. The prerequisite for the formation of functional membrane proteins was the existence of peptides containing 20–25 amino acid, which were sufficiently long to span a membrane. This length requirement is rather

modest, considering that functional water-soluble proteins need to be markedly larger. For example, the shortest, protobiologically relevant proteins contain 45 amino acids (Szostak *et al.*, 2001) and the simplest self-reproducing protein system consists of proteins built of 33 amino acids. (Lee *et al.*, 1997)

Many peptides are attracted to water-membrane or water-oil interfaces. Once at the interface, most nonpolar peptides spontaneously fold into α -helices. Peptides that contain both polar and nonpolar amino acids tend to adopt amphipathic structures, in which amino acid side chains are immersed in media of similar polarity. Whenever the sequence permits, peptides fold into amphipathic helices at interfaces. The formation of ordered, helical structures is primarily governed by the sequence of polar and nonpolar amino acids. Considering that specific identities of side chains is less important, the existence of helical peptides in interfacial, protocellular environments should not have been rare.

Helical peptides located parallel to the interface could insert into the membrane and adopt a transmembrane conformation. However, insertion of a single helix usually involves a positive free energy change, even for fully nonpolar peptides. The main reason why insertion is unfavorable is that polar groups in the peptide backbone and some side chains, which remain at least partially hydrated in water, become completely desolvated. The loss of solvation free energy is smaller for helices than for disordered structures because polar groups in the backbone are involved in intramolecular hydrogen bonding.

The unfavorable free energy of insertion can be regained by spontaneous association of peptides in the membrane into homomeric or heteromeric multimers. The first step in this process is the formation of dimers, although the most common structures involve aggregates of 4–7 helices. The helices could readily arrange themselves such that they form pores capable of transporting ions and small molecules across membranes. The stability of transmembrane aggregates of simple proteins is often marginal and, therefore, it can be regulated by environmental conditions, such as external electric fields or the specific nature of phospholipid headgroups, (Cafiso, 1994; Biggin and Sansom, 1996; Lear *et al.*, 1997; Tieleman *et al.*, 2001) or by small changes in the sequence of amino acids. (Fleming *et al.*, 1997; Fisher *et al.*, 1999) This ability to respond to environmental signals might have led to the earliest, although quite imprecise, regulation of transmembrane functions.

Clearly, a key step in the earliest evolution of integral membrane proteins was the emergence of selectivity for specific substrates. The selectivity of early channels was determined to some extent by all residues lining their lumen, which interact with substrates *via* electrostatic and van der Waals interactions. (Roux *et al.*, 2000) However, many contemporary simple channels employ filters or gates as the primary way to achieve selectivity. (Murata *et al.*, 2001; Roux and MacKinnon, 1999; Pinto *et al.*, 1997) From the evolutionary standpoint it is a very convenient solution because it requires placing just one or only a very few properly chosen residues in certain positions along the channel rather than imposing conditions on the whole sequence.

Many additional steps were required before simple aggregates of transmembrane peptides reached the structural and functional complexity, diversity and refinement of contemporary integral membrane proteins. The helices were connected by extra-membrane, hydrophilic linkers to stabilize them inside the membrane. The resulting, large proteins aggregated to even larger, higher-order structures. In many instances this step involved gene duplication. Protein sequences became optimized for highly specific functions. Perhaps most importantly, membrane proteins acquired large, water-soluble domains, which play a regulatory role or help to supply energy for active transport. This more advanced evolution of membrane proteins has been a subject of extensive studies. (Popot and Engelman, 2000) In the process, some intriguing connections between ion channels and enzymes have been uncovered. (Jan and Jan, 1992) The evolutionary history of membrane proteins is of special interest because it opened the doors for the emergence of multicellular

organisms endowed with nervous systems.

References

- Akeson, M., and Deamer, D. W. (1990). *Biochim. Biophys. Acta* **60**, 101–109.
- Allen, M. P., and Tildesley, D. J. (1987). “Computer Simulation of Liquids.” Oxford University Press, Oxford.
- Bassolino-Klimas, D., Alper, H. E., and Stouch, T. R. (1995). *J. Am. Chem. Soc.* **117**, 4118–4129.
- Berger, O., Edholm, O., and Jähnig, F. (1997). *Biophys. J.* **72**, 2002–2013.
- Berne, B. J., and Straub, J. E. (1997). *Curr. Opin. Struct. Biol.* **7**, 181–189.
- Bernstein, M. P., Dworkin, J. P., Sandford, S. A., and Allamandola, L. (2001). *Meteoritics and Planetary Science* **36**, 351–358.
- Biggin, P. C., and Sansom, M. S. (1996). *Biophys. Chem.* **60**, 99–110.
- Brack, A., and Barbier, B. (1990). *Origins Life Evol. Biosphere* **20**, 139–144.
- Cafiso, D. S. (1994). *Ann. Rev. Biophys. Biomol. Struct.* **23**, 141–165.
- Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, III, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Siebel, G. L., Singh, U. C., Wiener, P. K., and Kollman, P. A. (1999). AMBER 6. University of California, San Francisco.
- Chipot, C., and Pohorille, A. (1998). *J. Am. Chem. Soc.* **120**, 11912–11924.
- Chiu, S.-W., Clark, M., Balaji, V., Subramaniam, S., Scott, H. L., and Jakobsson, E. (1995). *Biophys. J.* **69**, 1230–1245.
- Corey, M. J., and Corey, E. (1996). *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11428–11434.
- Damodaran, K. V., Merz, Jr., K. M., and Gaber, B. P. (1995). *Biophys. J.* **69**, 1299–1308.
- Deamer, D. W. (1992). *Adv. Space Res.* **12**, 183–189.
- Deamer, D. W. (1997). *Microbiol. Mol. Bio. Rev.* **61**, 239–261.
- Deamer, D. W., and Nichols, J. W. (1989). *J. Membrane Biol.* **107**, 91–103.
- Deamer, D. W., and Pashley, R. M. (1989). *Origins Life Evol. Biosphere* **19**, 21–38.
- Duff, K. C., and Ashley, R. H. (1992). *Virology* **190**, 485–489.
- Essmann, U., Perera, L., and Berkowitz, M. L. (1995a). *Langmuir* **11**, 4519–4531.
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995b). *J. Chem. Phys.* **103**, 8577–8593.
- Feller, S. E. (2000). *Curr. Opin. Colloid Interface Sci.* **5**, 217–223.

- Fisher, L. E., Engelman, D. M., and Sturgis, J. N. (1999). *J. Mol. Biol.* **293**, 639–651.
- Fleming, K. G., Ackerman, A. L., and Engelman, D. M. (1997). *J. Mol. Biol.* **272**, 266–275.
- Forrest, L.R., and Sansom, M.S.P. (2000). *Curr. Opin. Struct. Biol.* **10**, 174–181.
- Frenkel, D., and Smit, B. (2001). “Understanding Molecular Simulations.” Academic Press, San Diego.
- Hager, A. J., Pollard, J. D., and Szostak, J. W. (1996). *Chemistry and Biology* **3**, 717–725.
- Hahn, K. W., Klis, W. A., and Stewart, J. M. (1990). *Science* **248**, 1544–1547.
- Jan, L. Y., and Jan, Y. N. (1992). *Cell* **69**, 715–718.
- Johnsson, K., Allemann, R. K., Widmer, H., and Benner, S. A. (1993). *Nature* **365**, 530–532.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). *J. Chem. Phys.* **79**, 926–935.
- Jorgensen, W. L., Madura, J. D., and Swenson, C. J. (1984). *J. Am. Chem. Soc.* **106**, 6638–6646.
- Joyce, G. F. (1996). *Current Biology* **6**, 965–967.
- Kollman, P. A. (1993). *Chem. Rev.* **93**, 2395.
- Lanyi, J. K. (1997). *J. Biol. Chem.* **272**, 31209–31212.
- Lanyi, J. K., and Pohorille, A. (2001). *Trends Biotechnol.* **19**, 140–144.
- Lear, J. D., Wasserman, Z. R., and DeGrado, W. F. (1988). *Science* **240**, 1177–1181.
- Lear, J. D., Schneider, J. P., Kienker, P. K., and DeGrado, W. F. (1997). *J. Am. Chem. Soc.* **119**, 3212–3217.
- Lee, D. H., Severin, K., Yokobayashi, Y., and Ghadiri, M. R. (1997). *Nature* **390**, 591–594.
- Lin, T. I., and Schroeder, C. (2001). *J. Virol.* **75**, 3647–3656.
- MacKerell, Jr., A. D., Brooks, B., Brooks, III, C. L., Nilsson, L., Roux, B., Won, Y., and Karplus, M. (1998). In “The Encyclopedia of Computational Chemistry” (*et al.* Schleyer, P. v. R., Ed.), Vol. 1, pp. 271–277. John Wiley and Sons, New York.
- Morowitz, H. J., Heinz, B., and Deamer, D. W. (1988). *Origins Life Evol. Biosphere* **18**, 281–287.
- Mould, J.A., Drury, J. E., Frings, S. M., Kaupp, U. B., Pekosz, A., Lamb, R. A., and Pinto, L. H. (2000). *J. Biol. Chem.* **275**, 31038–31050.
- Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymannh, J. B., Engel, A., and Fujiyoshi, Y. (2001). *Nature* **407**, 599–605.
- Oliver, A. E., and Deamer, D. W. (1994). *Biophys. J.* **66**, 1364–1379.
- Paula, S., Volkov, A. G., Van Hoeck, A. N., Haines, T. H., and Deamer, D. W. (1996). *Biophys. J.* **70**, 339–348.

- Paula, S., Volkov, A. G., and Deamer, D. W. (1998). *Biophys. J.* **74**, 319–327.
- Pearlman, D. A. (1994). *J. Phys. Chem.* **98**, 1487–1493.
- Perez-Paya, E., Houghton, R. A., and Blondelle, S. E. (1994). *Peptide Research* **7**, 286–288.
- Pinto, L. H., Holsinger, L. J., and Lamb, R. A. (1992). *Cell* **69**, 517–528.
- Pinto, L. H., Dieckmann, G. R., Gandhi, C. S., Papworth, C. G., Braman, J., Shaughnessy, M. A., Lear, J. D., Lamb, R. A., and DeGrado, W. F. (1997). *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11301–11306.
- Pohorille, A., and Wilson, M. A. (2001). *Cell. Mol. Biol. Lett.* **6**, 369–374.
- Pohorille, A., and Wilson, M.A. (1996). *J. Chem. Phys.* **104**, 3760–3773.
- Pohorille, A., Wilson, M.A., Chipot, C., M.H., New., and Schweighofer, K.S. (1999). In “Computational Molecular Biology” (J. Leszczynski, Ed.). Theoretical and Computational Chemistry, pp. 485–526. Elsevier, Amsterdam.
- Popot, J. L., and Engelman, D. M. (2000). *Annu. Rev. Biochem.* **69**, 881–922.
- Roberts, R. W., and Szostak, J. W. (1997). *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12297–12302.
- Roux, B., and MacKinnon, R. (1999). *Science* **424**, 100–102.
- Roux, B., Berneche, S., and Im, W. (2000). *Biochemistry* **44**, 13295–13306.
- Sakaguchi, T., Quiang, L., Pinto, L. H., and Lamb, R. A. (1997). *Proc. Natl. Acad. Sci. U. S. A.* **94**, 5000–5005.
- Sansom, M. S. P., Kerr, I. D., and Son, H. S. (1997). *Virology* **233**, 163–173.
- Schumaker, M. F., Pomes, R., and Roux, B. (2001). *Biophys. J.* **80**, 12–30.
- Schweighofer, K.J., and Pohorille, A. (2000). *Biophys. J.* **78**, 150–163.
- Severin, K., Lee, D. H., Kennan, A. J., and Ghadiri, M. R. (1997). *Nature* **389**, 706–709.
- Shen, L., Bassolino-Klimas, D., and Stouch, T. R. (1997). *Biophys. J.* **73**, 3–20.
- Shuck, K., Lamb, R. A., and Pinto, L. H. (2000). *J. Virol.* **74**, 7755–7761.
- Skulachev, V. P. (1984). In “Bioenergetics” (L. Ernster, Ed.). New Comprehensive Biochemistry, Vol. 23. Elsevier Science Publishers B. V., Amsterdam.
- Smondirev, A. M., and Berkowitz, M. L. (1999). *Biophys. J.* **76**, 2472–2478.
- Sun, K., and Mauzerall, D. (1996). *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10758–10762.
- Szostak, J. W., Luisi, P. L., and Bartel, D. P. (2001). *Nature* **409**, 387–390.
- Tieleman, D. P., Marrink, S. J., and Berendsen, H. J. C. (1997). *Biochim. Biophys. Acta* **1331**, 235–270.
- Tieleman, D.P., Berendsen, H.J.C., and Sansom, M.S.P. (2001). *Biophys. J.* **80**, 331–346.

- Wang, C., Takeuchi, K., Pinto, L., and Lamb, R. (1993). *J. Virology* **67**, 5585–5594.
- Wilson, M.A., and Pohorille, A. (1996). *J. Am. Chem. Soc.* **118**, 6580–6587.
- Wooley, G. A., and Wallace, B. A. (1992). *J. Memb. Biol.* **129**, 109–136.
- Zhong, Q., Hisslein, T., Moore, P. B., Newns, D. M., Pattnaik, P., and Klein, M. L. (1998a). *FEBS Lett.* **434**, 265–271.
- Zhong, Q., Jiang, Q., Moore, P. B., Newns, D. M., and Klein, M. L. (1998b). *Biophys. J.* **74**, 3–10.
- Zwanzig, R. W. (1954). *J. Chem. Phys.* **22**, 1420–1426.